Rayleigh-Ritz and Lanczos Methods for Symmetric Matrix Pencils*

Peter Lancaster
Department of Mathematics and Statistics
University of Calgary
Calgary, Alberta, T2N 1N4 Canada

and

Qiang Ye
Department of Applied Mathematics
University of Manitoba
Winnipeg, Manitoba
R3T 2N2 Canada

Submitted by Hans Schneider

ABSTRACT

We are concerned with eigenvalue problems for definite and indefinite symmetric matrix pencils. First, Rayleigh-Ritz methods are formulated and, using Krylov subspaces, a convergence analysis is presented for definite pencils. Second, generalized symmetric Lanczos algorithms are introduced as a special Rayleigh-Ritz method. In particular, an *a posteriori* convergence criterion is demonstrated by using residuals. Local convergence to real and nonreal eigenvalues is also discussed. Numerical examples concerning vibrations of damped cantilever beams are included.

1. INTRODUCTION

The classical variational methods for approximating eigenvalues numerically are best developed for self-adjoint problems $\lambda I - A$ where $A^* = A$, and for self-adjoint pencils $\lambda A - B$ where $A > 0$ (thus $A^* = A$, $B^* = B$, and A is positive definite). These statements hold whether the underlying

*Research supported in part by the National Sciences and Engineering Research Council of Canada.

655 Avenue of the Americas, New York, NY 10010 0024-3795/93/$6.00

space is finite- or infinite-dimensional. In the context of the design and analysis of algorithms the finite-dimensional case is paramount, of course, and the spectrum for either of these classical problems is confined to the real line.

In this paper some generalizations of variational methods are made for matrix pencils $\lambda A - B$ with $A^* = A$, $B^* = B$ and for which the condition $A > 0$ is dropped. However, (largely for convenience) it is assumed throughout that A is nonsingular. Thus A will be nonsingular and *indefinite*. It therefore generates a nondegenerate indefinite scalar product on $\mathbb{C}^n$, or $\mathbb{R}^n$, as the case may be. In this context an eigenvalue λ of $\lambda A - B$ is said to be of *positive type* (or *negative type*) if $x^*Ax > 0$ (or < 0, respectively) for all eigenvectors x associated with λ.

Starting with the two classical problems mentioned above, there is a hierarchy of symmetric problems of increasing generality and complexity. These include:

1. *Definite pencils*, which have the defining property that, for some α, $\beta \in \mathbb{R}$, $\alpha A + \beta B > 0$. It is necessary and sufficient that $A^{-1}B$ be diagonable and all eigenvalues be real and obey a *separation property*, i.e., either all the eigenvalues of positive type are greater than the eigenvalues of negative type or vice versa (see [11]). Clearly, the bilinear parameter transformation $\lambda = \alpha\mu(1 - \beta\mu)^{-1}$ reduces such a pencil to the classical form.
2. *Diagonable pencil with real spectrum.* A pencil $\lambda A - B$ is said to be diagonable if $A^{-1}B$ is similar to a diagonal matrix. A significant class of diagonable pencils with real spectrum has recently been investigated (see [2] and [3]). These include models of some conservative vibrating systems under the action of gyroscopic forces. In contrast to class 1, the "separation" property of the eigenvalues need not hold.
3. *General symmetric pencils* with $A^* = A$, $\det A \neq 0$, and $B^* = B$. In this case the spectrum is symmetric with respect to the real line, but generally the pencil is not diagonable and has nonreal eigenvalues.

For a detailed discussion of canonical structures for these problem classes see [6], for example. Note also that parameter shifts $\mu = \lambda - \lambda_0$ leading to pencils $\mu A - (B - \lambda_0 A)$ (and which are commonly used to shift an interesting part of the spectrum near to the origin) do not affect the problem classes defined here.

We briefly mention some relevant extensions of variational methods. First, minimax principles are known to extend to definite pencils. Earlier papers in this direction are by R. S. Phillips, B. Textorious, and G. W. Stewart. More recently, it has been shown that minimax (or maximin) characterizations of real eigenvalues can be extended to real eigenvalues of class-3 problems (see the review paper [12] for references and discussion). Rayleigh-quotient algorithms for the real eigenvalues follow readily and are investigated in [11] and [22]. Local convergence results for problems of class 3 and global convergence results for definite pencils are obtained.

In this paper we consider the development of Rayleigh-Ritz methods for the above problem classes with application to the theory behind the Lanczos tridiagonalization algorithm. An intriguing aspect of this problem area is the fact that the algorithm is known to "work" on problems of class 3, as demonstrated in the numerical examples given by Parlett and Chen [16], for example. The theoretical explanations of convergence properties remain to be developed. It is our objective to take some steps in that direction. In particular, the algorithms discussed are not new, and our interest here is not in algorithm developments but rather in the analysis of existing methods when applied to a wider class of problems. Discussions of the algorithmic aspect can be found in [5] and [16].

We implement the Rayleigh-Ritz process as follows (cf. [21], for example). Given $n \times n$ hermitian matrices A and B with A nonsingular and a subspace S_m of $\mathbb{C}^n$, choose a basis $p_1, \dots, p_m$ for S_m (thus, dim $S_m = m$ and $m \leqslant n$), and (using the standard inner product) form $m \times m$ matrices

$$A_1 = \left[(Ap_j, p_k) \right]_{j,k=1}^{m} \qquad B_1 = \left[(Bp_j, p_k) \right]_{j,k=1}^{m}.$$

We call $\lambda A_1 - B_1$ the *compression pencil* of $\lambda A - B$ with respect to $p_1, \dots, p_m$. Note that by defining the $n \times m$ matrix $P = [p_1, p_2, \dots, p_m]$ we may write

$$A_1 = P^*AP, \qquad B_1 = P^*BP.$$

One may imagine a sequence of subspaces of increasing dimension and the approximation of the eigenvalues of $\lambda A - B$ by those of the corresponding sequence of compression pencils. Candidates for this sequence of subspaces include the Krylov subspaces $\mathscr{K}^m(q)$ generated by a vector $q \in \mathbb{C}^n$ as follows:

$$\mathscr{K}^m(q) = \operatorname{span}\left\{ q, A^{-1}Bq, \dots, (A^{-1}B)^{m-1} q \right\}.$$

These subspaces are useful in analysis of Lanczos's algorithm. They have been used successfully in the analysis of the algorithm as applied to symmetric matrices, and lead to an *a priori* error bound, which ensures that a Krylov subspace with relatively low dimension yields a high-precision approximation of eigenpairs (see [15], for example). Furthermore, there is an *a posteriori* error bound that monitors the convergence of eigenvalues in the Lanczos process. For nonsymmetric matrices, Lanczos's algorithm has a generalized version (see [13]), but some difficulties arise in generalizing the classical analysis to this two-sided iteration (see [7] for example).

For problem classes 1, 2, and 3, we can take some advantage of the symmetric structure of pencils by introducing an *indefinite* inner product induced by A. In particular, there is a symmetric version of the Lanczos algorithm for hermitian pencils which can be analyzed in this way, and, as we shall see, it can also be regarded as a Rayleigh-Ritz method (in the above sense) using Krylov subspaces. In particular, it will be shown that for definite pencils, the *a priori* bounds for classical problems can be suitably extended.

In the case of definite pencils it may be argued that the bilinear transformation $\lambda = \alpha\mu(1 - \beta\mu)^{-1}$ referred to above should be employed to transform to a classical problem on which one can then use classical methods. The main reasons for not doing this are: (1) This device will not be available more generally. (2) There is no known easy method for finding α and β, and this is likely to be particularly delicate when the separation between the two eigenvalue types is not large (see examples at the end of this paper). (3) In numerical practice, the problem class under consideration is frequently unknown before computation begins.

In Section 2, we first introduce Rayleigh-Ritz methods for symmetric matrix pencils and, using Krylov subspaces, we give bounds for eigenvalue approximations generalizing the results of Kaniel, Paige, and Saad referred to above. Then, in Section 3, a bound based on Chebyshev polynomials is presented. After that, we formulate the generalized symmetric Lanczos algorithms as a special Rayleigh-Ritz method in Section 4. We discuss an *a posteriori* convergence criterion for both definite and indefinite pencils. For the indefinite case, instead of using residuals, we take advantage of the tridiagonal structure. In particular, by retaining symmetry, order-2 local approximations for real and nonreal eigenvalues are obtained in Section 6. The difficulty in obtaining eigenvalue bounds in terms of residuals is discussed in a more general setting by Kahan, Parlett, and Jiang [8]. They turn to backward error analysis for nonsymmetric problems and obtain first-order estimates for the norm of the (backward) perturbation matrix, and hence eigenvalue approximation, in terms of the norms of appropriate residuals. Finally, in Section 7, we present some numerical examples.

2. EIGENVALUE APPROXIMATION USING RAYLEIGH-RITZ METHODS

As above, let A, B be $n \times n$ hermitian matrices with A nonsingular (when we say that the pencil $\lambda A - B$ is nonsingular). If subspace $\mathcal{S}$ of $\mathbb{C}^n$

has a basis $p_1, p_2, \dots, p_m$, we may form the corresponding compression pencil $\lambda A_1 - B_1$, as in Section 1. This pencil obviously depends on the choice of basis for $\mathscr{S}$, but the spectrum of $\lambda A_1 - B_1$ is independent of this choice. If the basis is orthonormal, A_1, B_1 are said to be *orthogonal compressions* of A and B on $\mathscr{S}$, respectively.

It is possible that the compression A_1 is singular even though A is nonsingular. In this case, there is an $x \in \mathscr{S}$ such that $x^*Ay = 0$ for all $y \in \mathscr{S}$ (see [6]), and in particular the Lanczos algorithm may "break down" (see Section 4). In such a case, the compression pencil admits infinite eigenvalues though the original pencil does not. The appearance of very large eigenvalues when A_1 is "nearly" singular is also a problem. But even in this case some other eigenvalues of the compression pencil may provide good approximations. This is most easily seen in the special case of pencils $\lambda A - I$ or $\lambda^{-1}I - A$, where the appearance of zero eigenvalues does not affect the convergence of extreme eigenvalues.

If (θ, u) is an eigenpair of $\lambda A_1 - B_1$, i.e. $B_1 u = \theta A_1 u$ with $u \neq 0$, and $y = Pu \in \mathscr{S}$, we call (θ, y) a *Ritz pair* of the pencil $\lambda A - B$ on $\mathscr{S}$. It is easy to check that the pair has an orthogonality property, namely,

$$(\theta A - B)y \in \mathscr{S}^{\perp}, \tag{2.1}$$

where $\mathscr{S}^{\perp}$ denotes the orthogonal complement of $\mathscr{S}$ in $\mathbb{C}^n$. In this way, we seek an approximation y from a subspace $\mathscr{S}$ for which the residual $(\theta A - B)y$ is orthogonal to the same space $\mathscr{S}$.

Now we take the subspace $\mathscr{S}$ to be a Krylov space $\mathscr{K}^m(q)$. We say that $\mathscr{K}^m(q)$ is *A-nondegenerate*, if $A_1 = K^*AK$ is invertible, where $K = [q, (A^{-1}B)q, \dots, (A^{-1}B)^{m-1}q]$. In the sequel, we always consider A-nondegenerate Krylov subspaces. It is easy to see that

$$\mathscr{K}^m(q) = \{\pi(A^{-1}B)q : \pi \in \mathscr{P}^{m-1}\},$$

where $\mathscr{P}^{m-1}$ denotes the set of all polynomials of degree not exceeding $m-1$. In [9], [14], and [17], an approximation error bound for the Rayleigh-Ritz method using a Krylov subspace for a hermitian matrix was derived. The original analysis there depends on the classical minimax theorems. Here we extend the analysis to the definite-pencil case. The methods that we use are parallel to theirs and are based on the following generalized minimax theorem, which is proved in [11].

THEOREM 2.1. *Let $\lambda A - B$ be a definite nonsingular hermitian pencil, and suppose $\lambda_{r+1} \leqslant \cdots \leqslant \lambda_n < \lambda_1 \leqslant \cdots \leqslant \lambda_r$ are the eigenvalues and*

$x_{r+1}, \dots, x_n, x_1, \dots, x_r$ *are the corresponding eigenvectors, with* $\lambda_1, \dots, \lambda_r$ *of positive type and* $\lambda_{r+1}, \dots, \lambda_n$ *of negative type. Then for* $i = 1, 2, \dots, r$,

$$\lambda_i = \min_{\substack{x \in S_i \\ x^*Ax > 0}} \frac{x^*Bx}{x^*Ax} = \sup_{\dim S = n-i+1} \inf_{\substack{x \in S \\ x^*Ax > 0}} \frac{x^*Bx}{x^*Ax},$$

where $S_i = \operatorname{span}\{x_i, \dots, x_r, x_{r+1}, \dots, x_n\}$.

Two lemmas will also be helpful.

LEMMA 2.2. *Let* (θ, y) *be a Ritz pair of* $\lambda A - B$ *associated with an A-nondegenerate Krylov subspace* $\mathscr{K}^m(q)$ *for which* $y^*Ay \neq 0$. *Then*

$$\{x \in \mathscr{K}^m(q) : x^*Ay = 0\} = \{\pi(A^{-1}B)q : \pi \in \mathscr{P}^{m-1}, \pi(\theta) = 0\}.$$

Proof. Let $\mathscr{S}_1 = \{x \in \mathscr{K}^m(q) : x^*Ay = 0\}$ and $\mathscr{S}_2 = \{z = \pi(A^{-1}B)q : \pi \in \mathscr{P}^{m-1}, \pi(\theta) = 0\} \subset \mathscr{K}^m(q)$. Using the orthogonality property (2.1), it is easy to see that $y^*Ax = 0$ for any $x = (A^{-1}B - \theta I)x_1 \in \mathscr{S}_2$. Then $\mathscr{S}_2 \subset \mathscr{S}_1$. Also

$$\mathscr{S}_2 = \{z = \pi(A^{-1}B)(A^{-1}Bq - \theta q) : \pi \in \mathscr{P}^{m-2}\} = \mathscr{K}^{m-1}(A^{-1}Bq - \theta q).$$

Now $\mathscr{K}^m(q)$ is A-nondegenerate; therefore $(A^{-1}B)^{m-1}q \notin \mathscr{K}^{m-1}(q)$. This implies

$$(A^{-1}B)^{m-2}(A^{-1}Bq - \theta q) \notin \mathscr{K}^{m-2}(A^{-1}Bq - \theta q).$$

So $\dim \mathscr{K}^{m-1}(A^{-1}Bq - \theta q) = m - 1$; hence $\dim \mathscr{S}_2 = m - 1$. But $\dim \mathscr{S}_1 = m - 1$ because $y \notin \mathscr{S}_1$. Thus $\mathscr{S}_2 = \mathscr{S}_1$ and the lemma is proved. ■

Consider a definite pencil $\lambda A - B$. Obviously a compression pencil on $K^m(q)$ is also definite. Furthermore, the following property can be easily proved using the separation property of definite pencils and the inertia theorem.

LEMMA 2.3. *If* $\lambda A - B$ *is a nonsingular definite pencil and all eigenvalues of positive type are greater than all eigenvalues of negative type, then* $\lambda A_1 - B_1$, *the compression pencil with respect to an A-nondegenerate Krylov subspace, has the same properties. Furthermore, the number of positive (or negative) type eigenvalues of* $\lambda A_1 - B_1$ *cannot exceed the number of positive (or negative, respectively) type eigenvalues of* $\lambda A - B$.

Now, let $\lambda A - B$ be a nonsingular definite pencil. Let $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_r$ denote the eigenvalues of positive type and $\lambda_{r+1} \leqslant \cdots \leqslant \lambda_n$ denote the eigenvalues of negative type. By the separation property, we may assume $\lambda_n < \lambda_1$. With each eigenvalue λ_j we associate an eigenvector x_j, $j = 1, 2, \ldots, n$. By normalizing the eigenvectors we may then define

$$\delta_j = x_j^* A x_j = \begin{cases} 1, & 1 \leqslant j \leqslant r, \\ -1, & r+1 \leqslant j \leqslant n. \end{cases}$$

THEOREM 2.4. *Let $\mathscr{K}^m(q)$ be an A-nondegenerate Krylov subspace and $\lambda A_1 - B_1$ be the corresponding compression pencil of $\lambda A - B$. Let $\theta_1 \leqslant \cdots \leqslant \theta_p$ be the positive-type eigenvalues of $\lambda A_1 - B_1$. Then $p \leqslant r$.*

Let $q = \sum_{i=1}^n \alpha_i x_i$, and for any polynomial π define

$$a_k(\pi) = \sum_{i=1}^{n} |\pi(\lambda_i)\alpha_i|^2 (\lambda_i - \theta_1)^2 \cdots (\lambda_i - \theta_{k-1})^2 \delta_i$$

and

$$\epsilon_k = \min_{\substack{\pi \in \mathscr{P}^{m-k} \\ a_k(\pi) > 0}} \frac{\sum_{i=k+1}^n |\pi(\lambda_i)\alpha_i|^2 (\lambda_i - \theta_1)^2 \cdots (\lambda_i - \theta_{k-1})^2 (\lambda_i - \lambda_k)\delta_i}{a_k(\pi)}.$$

Then for $k = 1, \ldots, p$, we have

$$\lambda_k \leqslant \theta_k \leqslant \lambda_k + \epsilon_k.$$

Proof. Let $\{p_1, \ldots, p_m\}$ be a basis of $\mathscr{K}^m(q)$, $P = [p_1, \ldots, p_m]$, and $\lambda A_1 - B_1 = P^*(\lambda A - B)P$. We first show $\lambda_k \leqslant \theta_k$ for $k = 1, \ldots, p$. By Theorem 2.1

$$\lambda_k = \min_{\substack{x \in \mathscr{S}_k \\ x^*Ax > 0}} \frac{x^*Bx}{x^*Ax}$$

where $\mathscr{S}_k = \operatorname{span}\{x_k, \ldots, x_n\}$ with $\dim \mathscr{S}_k = n - k + 1$. Then $\dim[\mathscr{S}_k \cap \mathscr{K}^m(q)] \geqslant m - k + 1$. So there is a subspace $\tilde{\mathscr{S}}_k \subset \mathscr{S}_k \cap K^m(q)$ with dim $\tilde{\mathscr{S}}_k = m - k + 1$. Again by Lemma 2.3 and Theorem 2.1, for subspaces $\mathscr{T} \subset \mathbb{C}^m$ of dimension $m - k + 1$ and subspaces $\mathscr{S} \subset \mathscr{K}^m(q)$ $[\mathscr{S} = P(\mathscr{T})]$ of dimension $m - k + 1$,

$$\theta_k = \sup_{\dim \mathscr{T} = m-k+1} \inf_{\substack{z \in \mathscr{T} \\ z^* A_1 z > 0}} \frac{z^* B_1 z}{z^* A_1 z}$$

$$= \sup_{\dim \mathscr{T} = m-k+1} \inf_{\substack{z \in \mathscr{T} \\ z^* A_1 z > 0}} \frac{(Pz)^* B (Pz)}{(Pz)^* A (Pz)}$$

$$= \sup_{\substack{\dim \mathscr{S} = m-k+1 \\ \mathscr{S} \subset K^m(q)}} \inf_{\substack{x \in \mathscr{S} \\ x^* A x > 0}} \frac{x^* B x}{x^* A x}$$

$$\geqslant \inf_{x \in \tilde{\mathscr{S}}_k,\ x^* A x > 0} \frac{x^* B x}{x^* A x}$$

$$\geqslant \inf_{x \in \mathscr{S}_k,\ x^* A x > 0} \frac{x^* B x}{x^* A x}$$

$$= \lambda_k .$$

Now we establish the upper bound for θ_k. Let z_i be an eigenvector associated with θ_i, and let $y_i = Pz_i$, $\mathscr{T} = \{z \in \mathbb{C}^m : z^* A_1 z_i = 0,\ i = 1, \dots, k-1\}$, and $\mathscr{S} = \{y \in \mathscr{K}^m(q) : y^* A y_i = 0,\ i = 1, \dots, k-1\}$. Define

$$\Pi_1 = \Big\{ \pi \in \mathscr{P}^{m-1} : \pi(\theta_1) = \cdots = \pi(\theta_{k-1}) = 0,$$

$$\big[\pi(A^{-1}B)q\big]^* A \big[\pi(A^{-1}B)q\big] > 0 \Big\}$$

and

$$\Pi_2 = \big\{ \pi \in \mathscr{P}^{m-k} : a_k(\pi) > 0 \big\}.$$

It is easy to see from Theorem 2.1 that

$$\theta_k = \min_{z \in \mathscr{T},\ z^* A_1 z > 0} \frac{z^* B_1 z}{z^* A_1 z}.$$

Then using $\mathscr{S} = \{Pz : z \in \mathscr{T}\}$, we have

$$
\begin{aligned}
\theta_k &= \min_{y \in \mathscr{S},\, y^*Ay > 0} \frac{y^*By}{y^*Ay} \\
&= \min_{\pi \in \Pi_1} \frac{[\pi(A^{-1}B)q]^* B [\pi(A^{-1}B)q]}{[\pi(A^{-1}B)q]^* A [\pi(A^{-1}B)q]} \quad \text{(by Lemma 2.2)} \\
&= \min_{\pi \in \Pi_1} \frac{\sum_{i=1}^n \lambda_i |\alpha_i \pi(\lambda_i)|^2 \delta_i}{\sum_{i=1}^n |\alpha_i \pi(\lambda_i)|^2 \delta_i} \qquad \left(\text{using} \quad q = \sum_{i=1}^n \alpha_i x_i\right) \\
&= \min_{\pi \in \Pi_2} \frac{\sum_{i=1}^n \lambda_i |\alpha_i \pi(\lambda_i)|^2 (\lambda_i - \theta_1)^2 \cdots (\lambda_i - \theta_{k-1})^2 \delta_i}{\sum_{i=1}^n |\alpha_i \pi(\lambda_i)|^2 (\lambda_i - \theta_1)^2 \cdots (\lambda_i - \theta_{k-1})^2 \delta_i} \\
&= \lambda_k + \min_{\pi \in \Pi_2} \frac{\sum_{i=1}^n (\lambda_i - \lambda_k) |\alpha_i \pi(\lambda_i)|^2 (\lambda_i - \theta_1)^2 \cdots (\lambda_i - \theta_{k-1})^2 \delta_i}{\sum_{i=1}^n |\alpha_i \pi(\lambda_i)|^2 (\lambda_i - \theta_1)^2 \cdots (\lambda_i - \theta_{k-1})^2 \delta_i} \\
&\leqslant \lambda_k + \epsilon_k,
\end{aligned}
$$

where we note that $(\lambda_i - \lambda_k)\delta_i \leqslant 0$ for $1 \leqslant i \leqslant k$. ∎

We mention in passing that although the theorem is stated in the context of a Krylov subspace, the inequality $\lambda_k \leqslant \theta_k$ is true for general subspaces. Also there are dual results for negative-type eigenvalues or for the case where the positive-type eigenvalues are less than the negative-type eigenvalues. We omit these statements.

3. A COMPUTABLE ERROR BOUND

In this section, we use Theorem 2.4 to derive a computable error bound using some familiar polynomials.

Chebyshev polynomials have been widely used in numerical analysis. Their magnitudes are bounded by 1 in $[-1, 1]$, while increasing very rapidly outside the interval and as the degree increases. This nice property is used in some acceleration techniques. In particular, it can be used in Theorem 2.4 to give computable bounds. In our next discussion we seek some polynomials that have similar properties.

The approximation bound ϵ_k in Theorem 2.4 can be further bounded by choosing an appropriate polynomial in the minimization functional. Such a polynomial p should have the property that $|p(\lambda_i)/p(\lambda_k)|$ is small for $i > k$, or equivalently, $p(\lambda_k)$ is large while $p(\lambda_i)$ is bounded for $i > k$. Hence a polynomial which is bounded in $[\lambda_{r+1}, \lambda_n]$, $[\lambda_{k+1}, \lambda_r]$ but increases rapidly as λ increases from λ_n or decreases from λ_{k+1} will meet this requirement.

In general, precise solutions for this problem are complicated (see pp. 287–289 of [1]), so we will consider a relatively simple special case and suppose that the two intervals have the same length. For two general intervals it is always possible to expand the shorter one at the outer end to make two intervals of equal length, with the intermediate part fixed. However, this expansion will generally reduce the rate of increase of $p(x)$ in the intermediate part. Consider special intervals of $[-a, -1]$ and $[1, a]$ with $a > 1$. Generally two equal-length intervals can be transformed to this case by scaling and a shift of origin. In this case, the two intervals are symmetric with respect to the origin. We therefore consider polynomials which are symmetric about the origin (necessarily of even degree) and can be expressed in the form

$$p(x) = (x - a_1) \cdots (x - a_l)(x + a_1) \cdots (x + a_l)$$

$$= (x^2 - a_1^2) \cdots (x^2 - a_l^2),$$

where $a_i \geqslant 0$. Thus

$$p(x) = q(x^2)$$

for some polynomial q with degree l.

We want $p(x)$ to be bounded by 1 in $[-a, -1]$ and $[1, a]$ and to increase rapidly as x decreases from 1; equivalently, we need a $q(u)$ that is bounded by 1 in $[1, a^2]$ and increases rapidly as u decreases from 1. It is well known that the Chebyshev polynomials on $[1, a^2]$ have this property. Let T_l denote the Chebyshev polynomial of degree l with fundamental domain $[-1, 1]$. We take q to be the Chebyshev polynomial of degree l with interval $[1, a^2]$, i.e.

$$q(u) = T_l\left(\frac{2u - a^2 - 1}{a^2 - 1}\right).$$

Hence the corresponding p is given by $q(x^2)$. We denote

$$p_{l,a}(x) = T_l\left(\frac{2x^2 - a^2 - 1}{a^2 - 1}\right). \tag{3.1}$$

On p. 287 of [1] the problem of finding the polynomial that has the least deviation from zero in two intervals $[-a, -1]$ and $[1, a]$ among all monic polynomials of degree n is posed. For the even-degree case $n = 2l$, the polynomial is given by $L_{2l} p_{l,a}(x)$ with the minimal deviation $L_{2l} = (a^2 - 1)^l / 2^{2l-1}$. This is consistent with the polynomial $p_{l,a}(x)$ that we derived from another point of view. However, for the odd-degree case, the analytic solution becomes complicated, and we use the bounds obtained for the next lower (even) degree. Such polynomials are also discussed in [18] from the point of view of orthogonal polynomials.

A straightforward substitution of a polynomial of the form (3.1) into Theorem 2.4 yields:

COROLLARY 3.1. *Under the assumptions of Theorem* 2.4, *let*

$$c_k(x) = p_{l,a}\left(\frac{x - \alpha}{\beta}\right),$$

where $\alpha = (\lambda_{k+1} + \lambda_n)/2$, $\beta = (\lambda_{k+1} - \lambda_n)/2$, $l = [(m-k)/2]$, *and* $a = \max_{i \neq k} |\lambda_i - \alpha|/\beta$, *and let*

$$M_1 = \sum_{i=r+1}^{n} \frac{|\alpha_i|^2(\lambda_i - \theta_1)^2 \cdots (\lambda_i - \theta_{k-1})^2}{|\alpha_k|^2(\lambda_k - \theta_1)^2 \cdots (\lambda_k - \theta_{k-1})^2},$$

$$M_2 = \sum_{i=k+1}^{n} \frac{|\alpha_i|^2(\lambda_i - \theta_1)^2 \cdots (\lambda_i - \theta_{k-1})^2 |\lambda_i - \lambda_k|}{|\alpha_k|^2(\lambda_k - \theta_1)^2 \cdots (\lambda_k - \theta_{k-1})^2}.$$

Assuming $c_k(\lambda_k)^2 - M_1 > 0$, *we have*

$$\lambda_k \leqslant \theta_k \leqslant \lambda_k + \frac{M_2/c_k(\lambda_k)^2}{1 - M_1/c_k(\lambda_k)^2}. \tag{3.2}$$

To illustrate, the case $k = 1$ of (3.2) concerns the smallest eigenvalue of positive type, λ_1, and its approximation θ_1. In this case we have

$$M_1 = \sum_{i=r+1}^{n} \left|\frac{\alpha_i}{\alpha_1}\right|^2, \qquad M_2 = \sum_{i=2}^{n} \left|\frac{\alpha_i}{\alpha_1}\right|^2 |\lambda_i - \lambda_1|,$$

and (3.2) holds provided $c_1(\lambda_1)^2 > M_1$.

The bounds of (3.2) suggest that the interior extreme eigenvalues of the two types (i.e. λ_1, λ_2, etc.) are expected to converge first and the convergence rate depends on the gap between the two parts of the spectrum. In numerical experiments for Lanczos algorithms, however, the bigger and smaller eigenvalues of the whole spectrum (i.e. λ_r, λ_{r-1}, etc.) usually converge first, as in the classical case, and then, if the definite pencil has a good separation of the two eigenvalue types, the interior extreme eigenvalues emerge before the other eigenvalues converge (see Section 7).

We note that our results only bound the difference between $\theta_1, \dots, \theta_{p-1}, \theta_p$ and $\lambda_1, \dots, \lambda_{p-1}, \lambda_p$, but do not say anything about how θ_p, θ_{p-1} approach λ_r, λ_{r-1}. The convergence property of the latter is explained by a recent convergence analysis for nonsymmetric Lanczos algorithms [23], while that of the former may be a property peculiar to definite symmetric pencils.

4. LANCZOS ALGORITHMS

The Lanczos algorithm was originally introduced as a tridiagonalization method [13]. Later it was realized that it can be regarded as a Rayleigh-Ritz projection method using Krylov subspaces, and that it is an effective method to find some extreme eigenvalues of a large matrix. The tridiagonalization idea is naturally extended to nonsymmetric matrices to yield a two-sided Lanczos algorithm. In the remaining sections, we consider the Lanczos algorithm for symmetric matrix pencils—a special case of the two-sided algorithm.

We introduce the generalized symmetric Lanczos algorithms as a special Rayleigh-Ritz projection method. In the classical case of positive definite A, a sequence of A-orthogonal bases for Krylov subspaces is generated by the three-term recurrence

$$\beta_j q_{j+1} = (A^{-1}B)q_j - \beta_{j-1}q_{j-1} - \alpha_j q_j.$$

For $\lambda A - B$ with indefinite A, the same idea can be applied. Taking into account the change of signs of $q_j^*Aq_j$, we modify the recurrence as

$$\epsilon_{j+1}\beta_j q_{j+1} = (A^{-1}B)q_j - \epsilon_{j-1}\beta_{j-1}q_{j-1} - \epsilon_j\alpha_j q_j$$

(detailed in Algorithm 4.1). Then $\{q_1, \dots, q_j\}$ is an A-orthogonal basis of the Krylov subspace $\mathscr{K}^j(q_1)$. With this basis, the compression pencil has a simple form [as shown in Equation (4.2) below]; then the eigenvalues of the compression pencil are used to approximate the eigenvalues of the original pencil. More precisely, we formulate the generalized symmetric Lanczos algorithm for a hermitian pencil $\lambda A - B$ as in [12]. In effect, this is essentially the nonsymmetric Lanczos algorithm applied to $A^{-1}B$ with a special choice of initial vector.

ALGORITHM 4.1. Given an initial vector q_1 with $\epsilon_1 = q_1^*Aq_1 = \pm 1$, $\alpha_1 = q_1^*Bq_1$, the Lanczos algorithm generates $\{q_1, q_2, \dots, q_m\}$, $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$, $\{\beta_1, \beta_2, \dots, \beta_{m-1}\}$, and $\{\epsilon_1, \epsilon_2, \dots, \epsilon_m\}$ by the following recursion. Denote $\beta_0 q_0 = 0$ and $\epsilon_0 = 0$; then for $j = 1, 2, \dots$:

1. $r_j = (A^{-1}B)q_j - \epsilon_{j-1}\beta_{j-1}q_{j-1} - \epsilon_j\alpha_j q_j$;
2. if $r_j = 0$, stop; otherwise,
3. if $r_j^*Ar_j = 0$, stop; otherwise,
4. $\beta_j = \sqrt{|r_j^*Ar_j|}$, $\epsilon_{j+1} = \operatorname{sgn}(r_j^*Ar_j)$, $q_{j+1} = r_j/\beta_j\epsilon_{j+1}$, and $\alpha_{j+1} = q_{j+1}^*Bq_{j+1}$.

If there is no breakdown in generating $\{q_1, \dots, q_j\}$ (i.e., the "stop" in cases 2 and 3 does not occur), it is easy to check that for $Q_j = [q_1, \dots, q_j]$, we have

$$A^{-1}BQ_j = Q_jP_jT_j + r_je_j^*, \tag{4.1}$$

where $e_j^* = (0, \dots, 0, 1) \in \mathbb{R}^j$, $P_j = \operatorname{diag}(\epsilon_1, \dots, \epsilon_j)$, and

$$T_j = \begin{pmatrix} \alpha_1 & \beta_1 & 0 & \cdots & 0 \\ \beta_1 & \alpha_2 & \beta_2 & \cdots & 0 \\ 0 & \beta_2 & \alpha_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_j \end{pmatrix}.$$

Then, by induction, we can prove that $Q_j^* A Q_j = P_j$ and $Q_j^* A r_j = 0$. Thus

$$Q_j^* A Q_j = P_j, \qquad Q_j^* B Q_j = T_j. \tag{4.2}$$

In this algorithm, we normalize q_j so that $|q_j^* A q_j| = 1$. (See [16] for discussion of the alternative normalization $\|q_j\| = 1$.)

In Algorithm 4.1, after j steps have been performed, we obtain a pencil $\lambda P_j - T_j$. Then some familiar algorithm can be used to find the eigenvalues of $\lambda P_j - T_j$, which are used as an approximation of the eigenvalues (see [16]).

5. AN *A POSTERIORI* CONVERGENCE BOUND

In this section, we discuss an *a posteriori* convergence bound.

After step j of the Lanczos algorithm has been performed, we obtain a compression pencil $\lambda P_j - T_j = Q_j^*(\lambda A - B)Q_j$ as in (4.2). If θ_i and $s_i = (s_{1i}, \dots, s_{ji})^T$ are an eigenvalue and the corresponding eigenvector of $\lambda P_j - T_j$, then (θ_i, y_i) with $y_i = Q_j s_i$ is a Ritz pair and is used as an approximation of the eigenpairs of $\lambda A - B$. Multiplying (4.1) on the right by s_i, we obtain the following well-known identity concerning the residuals of Ritz pairs.

THEOREM 5.1. *If $\lambda P_j - T_j$ and r_j are generated by Algorithm 4.1 and $T_j s_i = \theta_i P_j s_i$ with $s_i = (s_{1i}, \dots, s_{ji})^T$, then, for $y_i = Q_j s_i$,*

$$By_i - \theta_i A y_i = s_{ji} A r_j. \tag{5.1}$$

Clearly, if the sequence $\{Ar_j\}$ is bounded, an s_{ji} of small absolute value demonstrates the convergence of the Ritz pair (θ_i, y_i) in the sense that the residual is small. Furthermore, in the classical case the error in approximation of the eigenvalues can be bounded by the residual (5.1) (see [8] or [15], for example). We show next that a suitably modified result also holds for definite pencils.

If $\lambda A - B$ is definite, it is clear that $\lambda P_j - T_j$ is definite. We then have a nonsingular matrix $S = (s_{kl})$ such that

$$S^* P_j S = P_j, \qquad S^* T_j S = P_j \Theta, \tag{5.2}$$

where $P_{sj} = \operatorname{diag}(\epsilon_1, \dots, \epsilon_j)$ with $\epsilon_j = \pm 1$, and $\Theta = \operatorname{diag}(\theta_1, \dots, \theta_j)$.

THEOREM 5.2. *Let $\{\lambda_i\}$ be the eigenvalues of a nonsingular definite pencil $\lambda A - B$, and let $S = (s_{kl})$ be the matrix satisfying (5.2). If (θ_i, y_i) is* *a Ritz pair, then*

$$\min_j \frac{(\lambda_j - \theta_i)^2}{\sqrt{1+\lambda_j^2}} \leqslant \frac{|s_{ji}|^2 \|Ar_j\|^2}{c(A,B)},$$

where $c(A, B) := \inf\{|x^*(A + iB)x| : \|x\| = 1\}$.

The number $c(A, B)$ is often known as the "Crawford number" of A and B (see [19], for example).

Proof. First, there exist $\alpha = \cos\theta$, $\beta = \sin\theta$ such that the smallest eigenvalue of $M = \alpha A + \beta B$ satisfies $\lambda_{\min}(M) = c(A, B) > 0$ (see [19]). Furthermore, there is a nonsingular X such that

$$X^*AX = P, \qquad X^*BX = PJ,$$

where $P = \operatorname{diag}(\epsilon_1, \dots, \epsilon_n)$ and $J = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. It is easy to see that $y_i^* A y_i = \epsilon_i$ and $\alpha\epsilon_i + \beta\epsilon_i\lambda_i > 0$. Let $z = (B - \theta_i A)y_i$, and, without loss of generality, assume that $B - \theta_i A$ is invertible. Then

$$\begin{aligned}
1 = |y_i A y_i| &= \left|z^*(B - \theta_i A)^{-1} A (B - \theta_i A)^{-1} z\right| \\
&= \left|z^* X (PJ - \theta_i P)^{-1} P (PJ - \theta_i P)^{-1} X^* z\right| \\
&= \left|z^* X \operatorname{diag}\left[\frac{\epsilon_j}{(\lambda_j - \theta_i)^2}\right] X^* z\right| \\
&= \left|z^* X \operatorname{diag}\left[\frac{\epsilon_j(\alpha\epsilon_j + \beta\epsilon_j\lambda_j)}{(\lambda_j - \theta_i)^2(\alpha\epsilon_j + \beta\epsilon_j\lambda_j)}\right] X^* z\right| \\
&\leqslant \max_j \frac{\alpha\epsilon_j + \beta\epsilon_j\lambda_j}{(\lambda_j - \theta_i)^2} \left|z^* X \operatorname{diag}\left[\frac{1}{\alpha\epsilon_j + \beta\epsilon_j\lambda_j}\right] X^* z\right| \\
&\leqslant \max_j \frac{\sqrt{1+\lambda_j^2}}{(\lambda_j - \theta_i)^2} |z^* M^{-1} z| \\
&\leqslant \max_j \frac{\sqrt{1+\lambda_j^2}}{(\lambda_j - \theta_i)^2} \frac{z^* z}{c(A,B)}.
\end{aligned}$$

Thus

$$\min_j \frac{(\lambda_j - \theta_i)^2}{\sqrt{1+\lambda_j^2}} \leqslant \frac{|s_{ji}|^2 \|Ar_j\|^2}{c(A,B)}. \qquad \blacksquare$$

6. APPROXIMATION OF EIGENVALUES IN CLASS 3

Theorem 5.2 applies to definite pencils, i.e. those of the first class described in the introduction. Here, a different technique is used which applies to simple eigenvalues of symmetric pencils of class 3, whether real or complex. As with Theorem 5.2, estimates are not practically computable, but they do give confidence that some convergence properties extend naturally to this case from the classical eigenvalue problem. This is also confirmed by our numerical experience.

Assume that Algorithm 4.1 is applied to a real $n \times n$ pencil of class 3 and, as is generally the case, runs through to completion in n steps without breakdown (see steps 2 and 3). Thus, a pencil $\lambda P - T$ is generated having the spectrum of $\lambda A - B$. Write $P = \mathrm{diag}[\epsilon_1, \epsilon_2, \ldots, \epsilon_n]$ and $P_j = \mathrm{diag}[\epsilon_1, \epsilon_2, \ldots, \epsilon_j]$, $\hat{P}_{n-j} = \mathrm{diag}[\epsilon_{j+1}, \ldots, \epsilon_n]$. Then write

$$T = \begin{bmatrix} T_j & \beta_j E \\ \beta_j E^T & \hat{T}_{n-j} \end{bmatrix},$$

where T_j is $j \times j$ (real, symmetric, and tridiagonal) and $E = e_j^T e_1$ has size $j \times (n-j)$. The number β_j, giving the magnitude of coupling between the compression $\lambda P_j - T_j$ and $\lambda P - T$, plays a prominent role in this analysis.

Let $K = P^{-1}T$, $K_j = P_j^{-1}T_j$, and make the partition

$$K = \begin{bmatrix} K_j & \epsilon_j \beta_j E \\ \epsilon_{j+1} \beta_j E^T & \hat{K}_{n-j} \end{bmatrix}. \tag{6.1}$$

We develop decompositions of K in two cases: when K_j has a real, and when it has a nonreal simple eigenvalue θ.

Case 1. Let θ be a simple *real* eigenvalue of K_j with real eigenvector x_1. As K_j is P_j-symmetric, we may assume that $x_1^T P_j$ is a corresponding left eigenvector with $x_1^T P_j x_1 = \nu_1$ $(= \pm 1)$. Let $x_1, x_2, \ldots, x_j$ be a P_j-orthogonal basis for $\mathbb{R}^j$, and let $\tilde{X} = [x_2 \; x_3 \; \cdots \; x_n]$, $X = [x_1 \; \tilde{X}]$ (see §I.1.2 of [6]). Then

$$X^T P_j X = \begin{bmatrix} \nu_1 & 0 \\ 0 & \tilde{P} \end{bmatrix},$$

where $\tilde{P} = \operatorname{diag}[\nu_2, \nu_3, \ldots, \nu_j]$ and $\nu_k = \pm 1$, $k = 2, 3, \ldots, n$. Consequently,

$$X^{-1} = \begin{bmatrix} \nu_1 x_1^T \\ \tilde{P}\,\tilde{X}^T \end{bmatrix} P_j, \tag{6.2}$$

and it can be verified that

$$X^{-1} K_j X = \begin{bmatrix} \theta & 0 \\ 0 & \tilde{K} \end{bmatrix}, \tag{6.3}$$

where $\tilde{K} = \tilde{P}\,\tilde{X}^T T_j \tilde{X}$.

Let $W = \operatorname{diag}[X, I_{n-j}]$, and we have

$$W^{-1} K W = \begin{bmatrix} X^{-1} K_j X & \epsilon_j \beta_j X^{-1} E \\ \epsilon_{j+1} \beta_j E^T X & \hat{K}_{n-j} \end{bmatrix}.$$

If the last row of X is denoted by $[\xi \;\; v^{\mathrm{T}}]$ ($\xi \in \mathbb{R}$), and we write $\sigma = \beta_j \xi$, then some manipulation using (6.1), (6.2), and (6.3) gives

$$W^{-1}(K)W = \left[\begin{array}{cc|cc} \theta & 0 & \nu_1 \sigma & 0 \\ 0 & \tilde{K} & \beta_j \tilde{P} v & 0 \\ \hline \epsilon_{j+1}\sigma & \epsilon_{j+1}\beta_j v^T & \multicolumn{2}{c}{\multirow{2}{*}{$\hat{K}_{n-j}$}} \\ 0 & 0 & & \end{array}\right]. \tag{6.4}$$

It is easily verified that if

$$H_1 := \begin{bmatrix} \tilde{P} & 0 \\ 0 & P_{n-j} \end{bmatrix}, \qquad L_1 = \left[\begin{array}{c|cc} \tilde{K} & \beta_j \tilde{P} v & 0 \\ \hline \epsilon_{j+1}\beta_j v^T & \multicolumn{2}{c}{\multirow{2}{*}{$\hat{K}_{n-j}$}} \\ 0 & & \end{array}\right], \tag{6.5}$$

then L_1 is H_1-symmetric (i.e., $H_1 L_1$ is real and symmetric).

Let e_j be the jth unit coordinate vector in $\mathbb{R}^{n-1}$. Using (6.4), we have

$$W^{-1}(\lambda_1 I - K)W = \begin{bmatrix} \lambda_1 - \theta & -\nu_1 \sigma e_j^T \\ -\epsilon_{j+1}\sigma e_j & \lambda_1 I - L_1 \end{bmatrix}. \tag{6.6}$$

Suppose that λ_1 is the eigenvalue of $\lambda P - T$ (and hence K) that is closest to θ, and that $\lambda_1 I - L_1$ is nonsingular. Then it follows from (6.6) (using a Schur complement) that

$$\lambda_1 - \theta = \nu_1 \epsilon_{j+1} \sigma^2 e_j^T (\lambda_1 I - L_1)^{-1} e_j,$$

or, defining

$$\kappa_1 = \nu_1 \epsilon_{j+1} e_j^T (\lambda_1 I - L_1)^{-1} e_j, \tag{6.7}$$

that

$$\lambda_1 - \theta = \sigma^2 \kappa_1. \tag{6.8}$$

Case 2. Let $\theta := \alpha + i\beta$ be a simple eigenvalue of K_j with $\alpha, \beta \in \mathbb{R}$, $\beta \neq 0$, and let x_1 be a corresponding eigenvector. Then $\bar{\theta}$ is also an eigenvalue of K_j with eigenvector $\bar{x}_1$. It is easily seen that $x_1^* P_j x_1 = 0$ and $x_1^* P_j \bar{x}_1 \neq 0$. So x_1 can be normalized so that $x_1^T P_j x_1 = 1$.

Now let $x_1, \bar{x}_1, x_3, \dots, x_j$ be a P_j-orthogonal basis for $\mathbb{C}^j$, and write

$$X = \begin{bmatrix} x_1 & \bar{x}_1 & x_3 & \cdots & x_j \end{bmatrix}, \qquad \tilde{X} = \begin{bmatrix} x_3 & \cdots & x_j \end{bmatrix}.$$

Thus,

$$X^* P_j X = \begin{bmatrix} P_0 & 0 \\ 0 & \tilde{P} \end{bmatrix}, \qquad X^{-1} K_j X = \operatorname{diag}\left[\theta, \bar{\theta}, \tilde{K}\ \right],$$

where

$$P_0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \tilde{P} = \tilde{X}^* P_j \tilde{X}, \quad \text{and} \quad \tilde{K} = \tilde{P}\, \tilde{X}^* K_j \tilde{X}.$$

As in case 1, write $W = \operatorname{diag}[X, I_{n-j}]$, but note that now

$$X^{-1} = \begin{bmatrix} \bar{x}_1^* \\ x_1^* \\ \tilde{P}\,\tilde{X}^* \end{bmatrix} P_j.$$

It is found that, if the last row of X is $[\xi, \bar{\xi}, u^*]$ and $\sigma = \beta_j \xi$, then [cf. (6.5) and (6.6)]

$$W^{-1}(K)W = \left[\begin{array}{ccc|cc} \theta & 0 & 0 & \sigma & 0 \\ 0 & \bar{\theta} & 0 & \bar{\sigma} & 0 \\ 0 & 0 & \tilde{K} & \beta_{j+1}\tilde{P}\,u & 0 \\ \hline \epsilon_{j+1}\sigma & \epsilon_{j+1}\bar{\sigma} & \epsilon_{j+1}\,\beta_j u^* & \multicolumn{2}{c}{\multirow{2}{*}{$\hat{K}_{n-j}$}} \\ 0 & 0 & 0 & & \end{array}\right]$$

$$= \left[\begin{array}{cc|c} \theta & 0 & \sigma e_{j-1}^* \\ 0 & \bar{\theta} & \bar{\sigma} e_{j-1}^* \\ \hline \epsilon_{j+1}\sigma e_{j-1} & \epsilon_{j+1}\bar{\sigma} e_{j-1} & L_2 \end{array}\right],$$

where e_{j-1} is the $(j-1)$st unit coordinate vector in $\mathbb{C}^{n-2}$, and we define

$$H_2 = \begin{bmatrix} \tilde{P} & 0 \\ 0 & P_{n-j} \end{bmatrix}, \qquad L_2 = \left[\begin{array}{c|cc} \tilde{K} & \beta_j \tilde{P}\,u & 0 \\ \hline \begin{matrix} \epsilon_{j+1}\,\beta_j u^* \\ 0 \end{matrix} & \multicolumn{2}{c}{\hat{K}_{n-j}} \end{array}\right] \tag{6.9}$$

[cf. (6.5)]. Note that now $H_2 L_2$ is hermitian.

Let λ_1 be the eigenvalue of K (or $\lambda P - T$) that is closest to θ, and we obtain

$$W^{-1}(\lambda_1 I - K)W$$

$$= \left[\begin{array}{cc|c} \lambda_1 - \theta & 0 & -\sigma e_{j-1}^* \\ 0 & \lambda_1 - \bar{\theta} & -\bar{\sigma} e_{j-1}^* \\ \hline -\epsilon_{j+1}\sigma e_{j-1} & -\epsilon_{j+1}\bar{\sigma} e_{j-1} & \lambda_1 I - L_2 \end{array}\right].$$

Assume that $\lambda_1 I - L_2$ is nonsingular, and define

$$\kappa_2 = \epsilon_{j+1} e_{j-1}^{*}(\lambda_1 I - L_2)^{-1} e_{j-1}. \tag{6.10}$$

Then the Schur-complement idea gives

$$\det\begin{bmatrix} \lambda_1 - \theta - \kappa_2 \sigma^2 & -\kappa_2 |\sigma|^2 \\ -\kappa_2 |\sigma|^2 & \lambda_1 - \bar{\theta} - \kappa_2 \bar{\sigma}^2 \end{bmatrix} = 0$$

and hence

$$\lambda_1 - \theta = \frac{\sigma^2 \kappa_2}{1 - \bar{\sigma}^2 \kappa_2 (\lambda_1 - \bar{\theta})^{-1}}, \tag{6.11}$$

the analogue of Equation (6.8).

We summarize the conclusions as follows:

PROPOSITION 6.1. *Let θ be a simple real (or nonreal) eigenvalue of the compression $\lambda P_j - T_j$ with right eigenvector x_1 satisfying $x_1^T P_j x_1 = \nu_1 = \pm 1$ (or $x_1^T P_j x_1 = 1$), and let $\sigma = \beta_j(e_j^T x_1)$. If λ_1 is the eigenvalue of $\lambda P - T$ closest to θ and is also simple, and if $\lambda_1 I - L_1$ (or $\lambda_1 I - L_2$) is nonsingular, then*

$$\lambda_1 - \theta = \begin{cases} \sigma^2 \kappa_1 & \text{if } \bar{\theta} = \theta, \\ \dfrac{\sigma^2 \kappa_2}{1 - \bar{\sigma}^2 \kappa_2 (\lambda_1 - \bar{\theta})^{-1}} & \text{if } \bar{\theta} \neq \theta, \end{cases} \tag{6.12}$$

where κ_1 and κ_2 are defined by (6.7) *and* (6.10), *respectively.*

To interpret this result, we must enquire about the behavior of κ_1 and κ_2 as $\sigma \to 0$. Observe that L_1 and L_2, and hence κ_1, κ_2, as well as σ, are functions of the initial vector q, chosen for the Lanczos algorithm. There is little or no control of the factor ξ $(= e_j^T x)$ in σ, as the normalization $x_1^T P_j x_1 = \pm 1$ does not control the magnitude of x_1 (unless $P_j = \pm I_j$). However, our numerical experience suggests that ξ is frequently small and may contribute significantly to the hypothesis that σ is small.

To get some quantitative estimates, let S be a matrix reducing the pair (H_1, L_1) of Equation (6.5) to the canonical form (H_c, J) (see Theorem I.3.3 of [6]). (For convenience, we present the case of real θ. The argument for nonreal θ is almost identical.) Thus,

$$S^{-1}L_1S = J, \qquad S^*H_1S = H_c, \tag{6.13}$$

where J is a Jordan canonical form and H_c is a canonical hermitian matrix satisfying (*inter alia*) $H_c^2 = I$. Then

$$S^{-1} = H_cS^*H_1$$

and

$$(\lambda_1 I - L_1)^{-1} = S(\lambda_1 I - J)^{-1}H_cS^*H_1.$$

Clearly, in the spectral norm, $\|H_c\| = \|H_1\| = 1$, so

$$\begin{aligned} \left\|(\lambda_1 I - L_1)^{-1}\right\| &\leqslant \|S\|^2\|\lambda_1 I - J\|^{-1} \\ &\leqslant \frac{\|S\|^2}{\min_{\mu\in\sigma(L_1)}|\lambda_1 - \mu|} \end{aligned} \tag{6.14}$$

if we assume that L_1 is diagonable.

Now let d be the separation of λ_1 from the other eigenvalues of $\lambda P - T$, i.e. the minimum of $|\lambda_1 - \beta|$ where β ranges over the eigenvalues of $\lambda P - T$ not equal to λ_1.

It might be expected that when σ is small, the eigenvalues of L_1 will be close to eigenvalues of κ other than λ_1. However, this may not be so in "nonclassical" cases. For example, if $a = \sqrt{(1-\sigma^2)(4-\sigma^4)}$, the matrix

$$\left[\begin{array}{c|cc} 0 & \sigma & 0 \\ \hline \sigma & \dfrac{2(1-\sigma^2)}{\sigma^2} & -\dfrac{a}{\sigma^2} \\ 0 & \dfrac{a}{\sigma^2} & -\dfrac{2}{\sigma^2} \end{array}\right]$$

has eigenvalues -2, -1, and 1, independent of σ. If $\sigma \to 0$, the (1, 1) entry does not converge to any of the eigenvalues. In considering asymptotic behavior as $\sigma \to 0$ we exclude such pathological cases by making the

CONTINUITY HYPOTHESIS. The initial vector q is varied in such a way that $\sigma \to 0$ and the eigenvalues of L_1 (or L_2) converge to eigenvalues of K other than λ_1.

With this hypothesis, we choose an eigenvalue $\mu_0(\sigma)$ of L_1 such that, for all sufficiently small σ,

$$\min_{\mu \in \sigma(L_1)} |\lambda_1 - \mu| = |\lambda_1 - \mu_0(\sigma)|.$$

Then $\mu_0(\sigma) \to \tilde{\lambda}$ as $\sigma \to 0$, where $\tilde{\lambda}$ is an eigenvalue of $\lambda P - T$ for which $|\lambda_1 - \tilde{\lambda}| = d$.

Now we may write, for μ an eigenvalue of L_1,

$$|\lambda_1 - \mu| \geqslant |\lambda_1 - \tilde{\lambda}| - |\tilde{\lambda} - \mu_0(\sigma)| = d - |\tilde{\lambda} - \mu_0(\sigma)|$$

and assume σ to be so small that

$$|\tilde{\lambda} - \mu_0(\sigma)| < \tfrac{1}{2}d. \tag{6.15}$$

Thus, $|\lambda_1 - \mu| \geqslant \frac{1}{2}d$, and we obtain from (6.14)

$$\|(\lambda_1 I - L_1)^{-1}\| \leqslant \frac{2\|S\|^2}{d}.$$

Combining this with (6.12) gives the following asymptotic result:

PROPOSITION 6.2. *Let the continuity hypothesis and the hypothesis of Proposition* 6.1 *hold. Suppose also that* L_1 *(or* L_2*) is diagonable. Then for all* *sufficiently small* $|\sigma|$,

$$|\lambda_1 - \theta| \leqslant \begin{cases} 2|\sigma|^2\|S\|^2/d & \text{if } \bar{\theta} = \theta, \\ \dfrac{2|\sigma|^2\|S\|^2}{d - 2|\sigma|^2\|S\|^2|\lambda_1 - \bar{\theta}|^{-1}} & \text{if } \bar{\theta} \neq \theta. \end{cases} \tag{6.16}$$

Clearly, L_1, L_2, and S are not within our computational grasp, so these bounds are of limited value. Nevertheless, $\|S\|$ will generally not be sensitive to $|\sigma|$. We might have anticipated sensitivity to small separation d of λ_1 from the rest of the eigenvalues of $\lambda P - T$, and to estimates θ with small imaginary parts. These effects are apparent in (6.16).

Note that, in the arguments of this section, it is the preservation of symmetry in Equations (6.5) and (6.9) that has permitted the derivation of the second relations in (6.12) and (6.16), and this also explains the appearance of $\|S\|^2$, rather than $\|S\|\,\|S^{-1}\|$ (the "condition" of S), in (6.16).

Let us conclude with the special case in which $A > 0$, whence $P = I$. In this case θ is necessarily real and it is well known that $|\lambda_1 - \theta| \leqslant \sigma$ (see [15], for example). Then the diagonability hypothesis and continuity hypothesis are unnecessary. Furthermore, in the canonical reduction of (6.13), $H_1 = H_C = I$ and S is unitary, and thus $\|S\| = 1$. Also, if $\sigma \leqslant \frac{1}{2}d$, (6.15) and hence (6.16) are true. Noting that $\sigma \leqslant 2\sigma^2/d$ if $\sigma \geqslant \frac{1}{2}d$, we have the following result. (Results of this kind are already known and can be found in [8] or [15]).

COROLLARY 6.3. *Assume that* $A > 0$. *Let* θ *be a simple eigenvalue of* $\lambda I - T_j$ *with right eigenvector* x_1 *satisfying* $\|x_1\|^2 = 1$, *and let* $\sigma = \beta_j(e_j^T x_1)$. *If* λ_1 *is the eigenvalue of* $\lambda I - T$ *closest to* θ *and is also simple, then* $\lambda_1 - \theta = \sigma^2\kappa_1$ *and* κ_1 *is given by* (6.7) *with* $\nu_1 = \epsilon_{j+1} = 1$. *Furthermore* $|\lambda_1 - \theta| \leqslant 2\sigma^2/d$ *for all* σ.

7. NUMERICAL EXAMPLES

We present some simple illustrative examples in this section. The first admits of comparison with the existing literature on Lanczos algorithms for symmetric pencils. The algorithm employed is an unsophisticated implementation of Algorithm 4.1 together with a full reorthogonalization process.

EXAMPLE 1. The following matrices A and B are used in [5]:

$$A = \begin{bmatrix} 4.3443 & -0.4696 & 0.1184 & 0.8482 & 2.3404 & -4.3342 \\ -0.4696 & -4.3485 & 6.4779 & 1.3731 & 0.1937 & -1.1516 \\ 0.1184 & 6.4779 & 2.7739 & 0.7720 & 0.8153 & 5.2281 \\ 0.8482 & 1.3731 & 0.7720 & 2.9806 & 0.6143 & -1.0325 \\ 2.3404 & 0.1937 & 0.8153 & 0.6143 & 1.9884 & -2.8104 \\ -4.3342 & -1.1516 & 5.2281 & -1.0325 & -2.8104 & 7.7535 \end{bmatrix},$$

$$B = \begin{bmatrix} 5.7850 & 8.1185 & 5.0490 & 1.7888 & 5.3932 & -0.7300 \\ 8.1185 & 23.2587 & -9.1936 & 1.8887 & 6.0846 & 1.9088 \\ 5.0490 & -9.1936 & 18.2674 & 5.1331 & -0.4157 & 0.4575 \\ 1.7888 & 1.8887 & 5.1331 & -1.2820 & 2.3567 & 1.4358 \\ 5.3932 & 6.0846 & -0.4157 & 2.3567 & 4.6237 & -3.8397 \\ -0.7300 & 1.9088 & 0.4575 & 1.4358 & -3.8397 & 5.9581 \end{bmatrix}.$$

The approximate spectrum of $\lambda A - B$, as determined by IMSL routines with an eight-digit word length, is

$$\{3.9995092, 1.9996290, -0.00041488241, -0.50014193,$$
$$-2.0001686, -2.9999571\},$$

with the first five of positive type and the last of negative type. Choosing the initial vector to be $(1, 1, \dots, 1)^*$, we run the program to the full six steps, and the following six Ritz values are generated:

$$\{3.9995105, 1.9996302, -0.00041206181, -0.50014337,$$
$$-2.0001669, -2.9999635\}.$$

EXAMPLE 2. Our main suite of examples comes from the analysis of vibrating, damped, cantilever beams in mechanical engineering. The inclusion of two forms of damping leads to the following quadratic eigenvalue problem (see [4], for example):

$$\lambda^2 p(x)u(x) + \lambda\left[\alpha u^{(4)}(x) + \gamma u(x)\right] + u^{(4)}(x) = 0 \tag{7.1}$$

on [0, 1], with boundary conditions

$$u(0) = u'(0) = 0, \qquad u^{(2)}(1) = u^{(3)}(1) = 0.$$

Here the beam itself is uniform (with respect to x), but the function $\rho(x)$ admits a nonuniform loading. The term $\gamma u(x)$ represents external viscous damping, and $\alpha u^{(4)}(x)$ represents internal damping of the beam. To solve (1) numerically, we use finite-element methods (and uniform hermite cubic elements) to discretize (7.1) and obtain the matrix problem

$$\lambda^2 M_n u + \lambda\left(\alpha n^2 K_n + \frac{\gamma J_n}{n^2}\right)u + K_n u = 0, \tag{7.2}$$

where n is the number of elements. We reformulate this as the "equivalent" linear pencil problem with

$$A = \begin{bmatrix} -K_n & 0 \\ 0 & M_n \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \alpha n^2 K_n + \gamma J_n/n^2 & M_n \\ M_n & 0 \end{bmatrix} \tag{7.3}$$

(see [6], for example). Note that if the beam has no loading, i.e., $\rho(x) = 1$, then there are only two different operators in the coefficients of (7.1). Thus $M_n = J_n$, and J_n, K_n can be simultaneously diagonalized. Thus the reduction to (7.3) is unnecessary. In this example, we therefore consider a loaded beam with

$$\rho(x) = \begin{cases} 1, & 0 \leqslant x < 0.5, \\ 2, & 0.5 < x \leqslant 1, \end{cases}$$

and various values of the damping parameters α and γ. Twenty elements are used, so that the matrix coefficients in (7.2) are of size 40 and A, B in (7.3) are of size 80. Computations are started with random vectors. In these examples, the convergence criterion of Section 6 is used to select the four best eigenvalue approximations after ten Lanczos steps. This is used in an attempt to simulate the computational practice of determining just a few eigenvalues by the Lanczos process. A trend toward the determination of the extreme eigenvalues and the interior extreme eigenvalues of the two types is immediately apparent.

In Figures 1 and 2, the eigenvalues of positive type, eigenvalues of negative type, and nonreal eigenvalues obtained by the QZ algorithm are marked by $\triangledown$ $\vartriangle$ and $\varnothing$, respectively, and those obtained from the Lanczos algorithm are marked by $\uparrow$, $\downarrow$, and $\times$, respectively. The horizontal arrows indicate the location of accumulation points for the eigenvalues of (7.1). Results for cases (a)–(d) are also presented in Table 1.

(a) $\alpha = 0.8$, $\gamma = 0.1$. In this case, (7.3) is definite with moderate separation of the two eigenvalue types. See Figure 1(a).

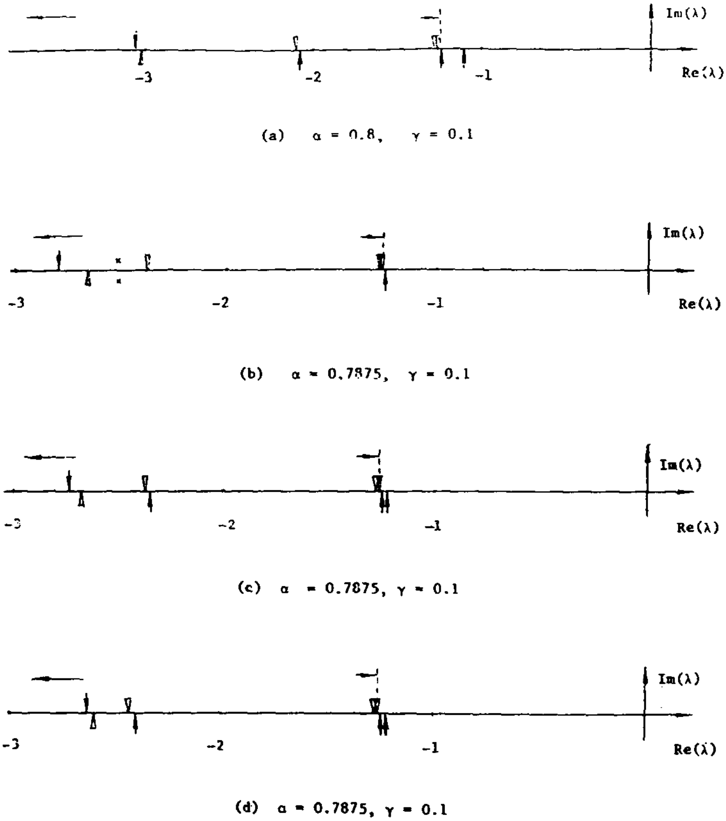

FIG. 1. Four computed eigenvalues for definite systems. $\triangledown$, $\vartriangle$, $\varnothing$ from QZ algorithm; $\uparrow$, $\downarrow$, $\times$ from Lanczos algorithm.)

(b) $\alpha = 0.7875$, $\gamma = 0.1$. In this case, (7.3) is definite, with small separation of the two eigenvalue types. Numerical errors lead to the appearance of a pair of conjugate eigenvalues with a small imaginary part, and the real part is stabilized at the mean of the two interior extreme eigenvalues. See Figure 1(b).

(c) *The same case as* (b) *but with a different starting vector.* This time the complex pairs do not appear. See Figure 1(c).

(d) $\alpha = 0.7865$, $\gamma = 0.1$. Again (7.3) is definite, with small separation of the two eigenvalue types. One set of results is indicated in Figure 1(d). With a different starting vector, a conjugate pair appears at steps 14, 16, 18 and then disappears, i.e., the pair does not stabilize.

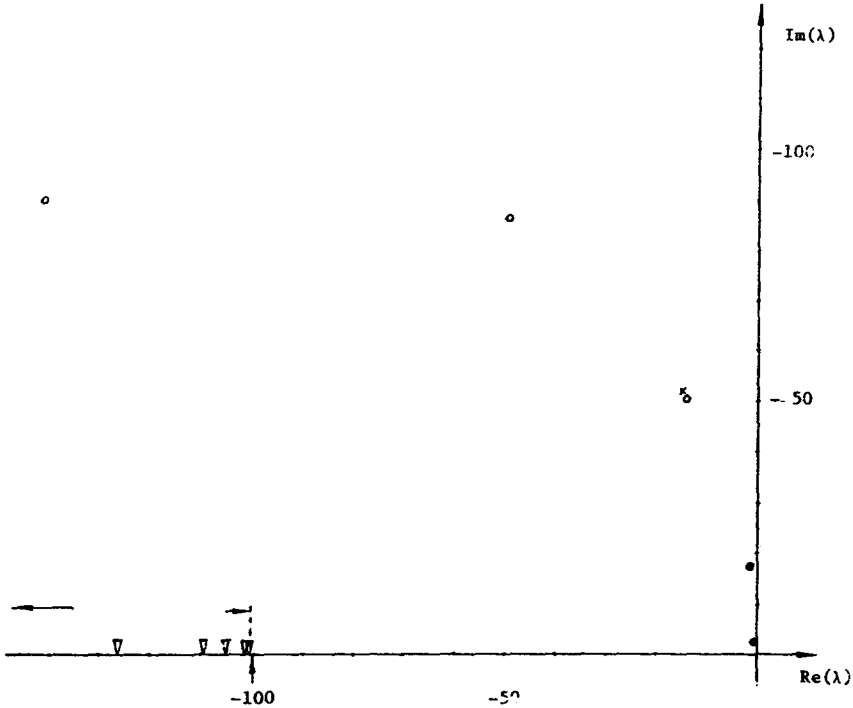

FIG. 2. Four computed eigenvalues for a mixed system of size 80 × 80. (▽, △, ∅ from QZ algorithm; ↑, ↓, × from Lanczos algorithm.)

TABLE 1
FOUR COMPUTED EIGENVALUES FOR A DEFINITE SYSTEM

		$\alpha = 0.8$, $\gamma = 0.1$		
QZ	-1.2500000	-1.2559522	-2.1080604	-3.0073822
Lanczos (a)	-1.11583692	-1.2499897	-2.0738695	-3.0587931
		$\alpha = 0.7875$, $\gamma = 0.1$		
QZ	-1.2698412	-1.2762257	-2.3753880	-2.6687567
Lanczos (b)	-0.84200535	-1.2698533	-2.5153135	
			$\pm 0.064878133i$	
Lanczos (c)	-1.2698457	-1.2747137	-2.4042470	-2.6356385
		$\alpha = 0.7865$, $\gamma = 0.1$		
QZ	-1.2714559	-1.2778075	-2.4348756	-2.6032984
Lanczos (d)	-1.2542141	-1.2714130	-2.4075413	-2.6326534

TABLE 2
COMPUTED EIGENVALUES FOR A MIXED SYSTEM

QZ	Lanczos
$-0.057325065 \pm 2.5169976i$	$-0.057330587 \pm 2.5174472i$
$-1.5846702 \pm 17.551290i$	$-1.5846670 \pm 17.551752i$
$-13.797698 \pm 50.610795i$	$-13.831803 \pm 50.592159i$
-100.00348	-100.09559

(e) $\alpha = 0.01$, $\gamma = 0.1$. In this case, (7.3) is not definite and there are five conjugate pairs of nonreal eigenvalues. However, the real eigenvalues of the two types of are still separated. Counting a conjugate pair as one eigenvalue, the four best approximations after ten steps are illustrated in Figure 2 and Table 2.

The results may be described as interesting, consistent with theory, and disappointing in accuracy, and therefore worthy of further investigation.

REFERENCES

1 N. Achieser, *Theory of Approximation*, Frederick Ungar, New York, 1956.
2 L. Barkwell and P. Lancaster, Overdamped and gyroscopic vibrating systems, *J. Appl. Mech.*, 59:176–181 (1992).
3 L. Barkwell, P. Lancaster, and A. Marcus, Gyroscopically stabilized systems: A class of quadratic eigenvalue problems with real spectrum, *Canad. J. Math.*, 44:42–53 (1992).
4 W. E. Baker, W. Woodam, and D. Young, Air and internal damping of thin cantilever beams, *Internat. J. Mech. Sci.* 9:743–766 (1967).
5 T. Ericsson and A. Ruhe, Lanczos algorithm and field of value rotations for symmetric matrix pencils, *Linear Algebra Appl.* 88/89:733–746 (1987).
6 I. Gohberg, P. Lancaster, and L. Rodman, *Matrices and Indefinite Scalar Products*, Birkhäuser, Basel, 1983.
7 G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins U.P., Baltimore, 1983.
8 W. Kahan, B. Parlett, and E. Jiang, Residual bounds on approximate eigensystems of nonnormal matrices, *SIAM J. Numer. Anal.* 19:470–484 (1982).
9 S. Kaniel, Estimates for some computational techniques in linear algebra, *Math. Comp.* 20:369–378 (1966).
10 P. Lancaster and M. Tismentsky, *The Theory of Matrices*, Academic, Orlando, 1985.
11 P. Lancaster and Q. Ye, Variational properties and Rayleigh quotient algorithms for symmetric matrix pencils, in *The Gohberg Anniversary Collection*, *Oper. Theory. Adv. Appl.* 40, Birkhäuser, Basel, 1989, pp. 247–278.

12 P. Lancaster and Q. Ye, Variational and numerical methods for symmetric matrix pencils, *Bull Austral. Math. Soc.* 43:1–17 (1991).
13 C. Lanczos, An iterative method for the solution of the eigenvalue problems of linear differential and integral operators, *J. Res. Nat. Bur. Standards Ser. B* 45:225–280 (1950).
14 C. Paige, The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices, Ph.D. Dissertation, Univ. of London, 1971.
15 B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
16 B. N. Parlett and H. C. Chen, Use of an indefinite inner product for computing damped natural modes, *Linear Algebra Appl.* 140:53–88 (1990).
17 Y. Saad, On the rate of convergence of the Lanczos and block Lanczos methods, *SIAM J. Numer. Anal.* 17:687–706 (1980).
18 Y. Saad, Iterative solutions of indefinite symmetric systems by methods using orthogonal polynomials over two disjoint intervals, *SIAM J. Numer. Anal.* 20:784–811 (1983).
19 G. W. Stewart, Perturbation bounds for the definite generalized eigenvalue problem, *Linear Algebra Appl.* 23:69–85 (1979).
20 G. Strang and G. J. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, N.J., 1973.
21 H. F. Weinberger *Variational Methods for Eigenvalue Approximation* SIAM, Philadelphia, 1974.
22 Q. Ye, Variational Principles and Numerical Algorithms for Symmetric Matrix Pencils, Ph.D. Thesis, Univ. of Calgary, 1989.
23 Q. Ye, A convergence analysis of nonsymmetric Lanczos algorithms, *Math. Comp.* 56:677–691 (1991).

Received 4 March 1992; final manuscript accepted 28 May 1992